DOCUMENT RESUME

ED 038 380 SP 003 853

AUTHOR	Medley, Donald M.; Hill, Russell A.
TITLE	Cognitive Factors in Teaching Style.
PUB DATE	70
NOTE	18p.; Paper presented at the AERA meeting, Minneapolis, March 1970
EDRS PRICE	EDRS Price MF-$0.25 HC-$1.00
DESCRIPTORS	*Cognitive Tests, *Internship Programs, Statistical Data, *Teacher Attitudes, Teacher Behavior, *Teacher Evaluation, *Teaching Experience, Teaching Styles
IDENTIFIERS	Flanders Interaction Analysis, OScAR 4V

ABSTRACT

A study of the relationship between teacher knowledge and teaching style was made in an eastern metropolitan area, with 53 first-year intern teachers in secondary schools whose answers in the Common Examination of the National Teacher Examination were available for analysis. Each teacher was visited in the classroom on four occasions by two observers, one trained in the Flanders system of Interaction Analysis and the other in the OScAR 4V system. The resulting observations and the 345 multiple-choice examination items were analyzed and evaluated, and the results are set out in four tables. Although the findings are tentative, they indicate a correlation between lecturing behavior and performance in the NTE examination, with science-oriented teachers lecturing more and listening less, while those with high scores in literature and the history and philosophy of education and teaching practices favor a dialog approach. The study suggests a close connection between the teacher's knowledge of teaching principles and practices and teaching style and, if based on larger numbers, could probably have provided additional data on the dynamics and etiology of teaching styles. (MBM)

ED038380

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

COGNITIVE FACTORS IN TEACHING STYLE

Donald M. Medley
Educational Testing Service
Princeton, N. J.

Russell A. Hill
Research for Better Schools
Philadelphia, Pa.

A paper presented at the Annual Meeting
of the American Educational Research Association
Minneapolis, Minnesota, March 4, 1970

SP003853

Table 1

Analysis of the Performance of the
91 Teachers on 19 Subtests of the
Common Examinations of the National Teacher Examinations

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square Obtained	Mean Square Expected
Candidates	90	253.20	2.81	$345\sigma^2_c + \sigma^2$
Subtests	18	348.64	19.37	$91\bar{K}\Theta^2_s + 91\sigma^2_i + \sigma^2$
Items (in subtests)	326	1,202.18	3.69	$91\sigma^2_i + \sigma^2$
Interaction, Candidate by Subtest	1,620	528.49	0.33	$\bar{K}\sigma^2_{cs} + \sigma^2$
Residual	29,340	5,089.91	0.17	σ^2
Total Variation	31,394	7,422.41		

$$\bar{K} = \frac{1}{18}\left(345 - \frac{\Sigma K^2_i}{345}\right)$$

K_i = number of items on subtest i

i = 1, 2, ..., 19

Table 2

Pooled Analysis of Variance
of Scores of 91 Teachers on 345 Items of the
Common Examinations of the National Teacher Examinations

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares
Candidates	90	253.20	2.81
General Knowledge (vs. Professional)	1	96.30	96.30
Science and Mathematics (vs. rest of General Knowledge)	1	17.07	17.07
Science (vs. Mathematics)	1	18.46	18.46
English (vs. Social Studies and Fine Arts)	1	45.91	45.91
Literature (vs. English)	1	35.46	35.46
Social Studies (vs. Fine Arts)	1	16.32	16.32
Foundations (vs. Teaching Principles and Practices)	1	27.61	27.61
History and Philosophy (vs. School and Society)	1	48.84	48.84
Items (within Subsets)	336	1,244.85	3.71
Candidate x General (vs. Professional Knowledge)	90	57.29	0.64
Candidate x Science and Mathematics (vs. Other General Knowledge)	90	78.17	0.87
Candidate x Science (vs. Mathematics)	90	42.62	0.47
Candidate x English (vs. Social Studies and Fine Arts)	90	41.15	0.46
Candidate x Literature (vs. English Mechanics and Effectiveness)	90	53.44	0.59
Candidate x English A (vs. English B)	90	23.00	0.26
Candidate x Social Studies (vs. Fine Arts)	90	44.42	0.49
Candidate x Foundations (vs. Teaching Principles and Practices)	90	20.39	0.23
Candidate x History and Philosophy (vs. School and Society)	90	22.18	0.25
Pooled Residual	30,150	5,235.73	0.17
Totals	31,394	7,422.41	

Table 3

Subtests Scores on the Common Examinations and
Contrasts for which the Null Hypothesis Was Rejected

Subtest	Number of Items	Mean Reliability Per Item	Contrasts
Science	30	.16	+ + + 0 0 0 0 0 0
Mathematics	20	.20	+ + - 0 0 0 0 0 0
English A	30	.11	+ - 0 + - + 0 0 0
English B	25	.09	+ - 0 + - - 0 0 0
Literature	19	.21	+ - 0 + + 0 0 0 0
Social Studies	30	.08	+ - 0 - 0 0 + 0 0
Fine Arts	16	.18	+ - 0 - 0 0 - 0 0
Teaching Principles and Practices	60	.03	- 0 0 0 0 0 0 - 0
History and Philosophy of Education	20	.03	- 0 0 0 0 0 0 + +
School and Society	22	.02	- 0 0 0 0 0 0 + -
Psychological Foundations plus Teacher Role plus School Organization	73	.04	- 0 0 0 0 0 0 + 0
Total	345		

Table 4

Regressions of Teacher Behaviors
-on National Teacher Examinations Common Examinations

Dependent Variables	Multiple R	Beta Weights on Test Subscores										
		Science	Mathematics	English A	English B	Literature	Social Studies	Fine Arts	Teaching Principles	History and Philosphy	School and Society	Psychology etc.
1. Lecturing Behavior (FIAT)	.66*	+.38*	+.13	+.13	+.18	-.37	-.15	+.16	-.42*	+.30	+.02	-.04
2. Lecturing Behavior (OScAR)	.56	+.14	+.05	+.19	+.19	-.30	+.30	+.20	-.42*	-.05	+.15	-.25
3. Modified Content Cross (FIAT)	.60*	+.39*	-.21	+.31	+.05	-.17	-.01	+.17	-.12	+.25	-.14	+.02
4. Criticizing Behavior (FIAT)	.55	-.02	-.12	-.48*	+.19	-.30	-.09	-.16	-.09	+.33	+.38	+.11
5. Rebuking Behavior (OScAR)	.48	-.06	-.18	-.27	+.26	-.52*	-.07	+.07	+.01	+.47	+.33	+.02
6. I-D Contrast on Student Response (FIAT)	.53	+.36	-.28	+.16	+.19	-.11	-.10	+.08	-.03	+.42*	+.00	-.18
7. Questioning Style (OScAR)	.53	-.04	+.27	-.19	-.33	+.20	-.09	-.08	-.27	-.09	+.46*	-.12
8. Listening Behavior (OScAR)	.48	-.19	+.25	-.05	-.46*	+.36	+.11	+.07	+.21	-.09	-.26	+.10

*P <.05

COGNITIVE FACTORS IN TEACHING STYLE

The findings that I am going to describe to you today were a byproduct of a larger study involving 70 first-year intern teachers in a large metropolitan area in the eastern United States. These 70 teachers were visited in their classrooms four times each by a pair of trained observers, and their behavior was recorded for not quite half an hour on each visit. Fifty-three of the 70 teachers had also taken the Common Examinations of the National Teacher Examinations just before beginning their first year of teaching, and we were able to retrieve their answer sheets for use in the present analysis. These 53 secondary school teachers, on whom both behavior records and test data were available, constitute the subjects with whom we are concerned today. The group included teachers of all four major subjects--science, mathematics, English, and social studies--at both the junior and senior high school level.

One of the two observers who visited each teacher was trained in the system of Interaction Analysis developed by Flanders (Amidon & Flanders, 1963)

and recorded verbal behavior according to that system. The other was trained to use a different technique, OScAR 4V (Medley, Impelletteri, & Smith, 1966), and recorded the same verbal behaviors using that system.

All observations were intercorrelated and submitted to a principal components analysis on the basis of which 15 scoring keys were built, 8 for OScAR, and 7 for the Flanders' system (Medley, & Hill, 1968, 1969). Scores on these 15 keys accounted for about two-thirds of all of the variance in the observations. These scores constituted the measures of teacher behavior, or style, used in the present study.

The form of the Common Examinations of the National Teacher Examinations taken by the 53 teachers contained 345 multiple-choice items. The 345 items were written according to a table of specifications which called for items representing 19 different content areas. About half of the items were designed to sample the teacher's knowledge of subject-matter content commonly included in secondary school curricula, such as science, mathematics, English, etc. The other half were designed to measure knowledge of the content of professional

education courses--history and philosophy of education, teaching principles and practices, and so on. Subscores were obtained for each teacher on each of 19 "subtests" made up of items from one of the 19 content areas.

Answer sheets were available for 91 teachers, including the 53 who were observed plus 38 others enrolled in the same program but not observed in their classrooms. These 91 papers were submitted to an analysis of variance of the form originally suggested by Hoyt (1941), to study the internal structure of the test. Hoyt used a two-way design without replication, items by candidates; in this case, there were 90 degrees of freedom for candidates, 344 for items, and 30,960 for error. We extended Hoyt's design by partitioning the 344 degrees of freedom for items into two portions. One portion, with 18 degrees of freedom, estimated between-items variance from variation between items on <u>different</u> subtests only; the other, with 326 degrees of freedom, estimated the same variance by comparing only items on the <u>same</u> subtest.

If you will consult Table 1 on your handout, you will note that the

---

Insert Table 1 about here

---

sum of squares for error was also partitioned in an analogous fashion. One portion, with 1,620 degrees of freedom, estimated errors of measurement from interaction between candidates and items on <u>different</u> subtests, and the other, with 29,340 degrees of freedom, estimated errors of measurement from interaction between candidates and items on the <u>same</u> subtest.

The fact that the former mean square is larger than the latter indicate that σ^2_{cs} is greater than zero; that is, that there is an interaction between candidates and subtests. Therefore we may not assume that all of the 19 subtests are measuring the same function since some candidates tend to do better on one subtest than they do on others. Or, to put it differently, since the rank order of true scores of the 91 candidates varies from one subtest to another, we must conclude that the subtests measure different functions. The practical interpretation is that there is information in the subtest scores

that does not appear in the total score, so we must retain at least some subtest scores for further analysis.

The question remains: how many, and which subtest scores should we retain? To answer this question we partitioned the 18 degrees of freedom between subtests, and the 1,620 for candidate by subtest interaction, into 18 parts each. In the case of subtests, there was one degree of freedom for each part, in the case of interaction, there was 90 degrees of freedom for each portion.

In making these partitions we used 18 orthogonal contrasts among subtests, reflecting 18 a priori hypotheses about how the content areas sampled by the 19 subtests might differ. The null hypothesis was rejected in eight instances and accepted in ten as regards between subtests variation. The null hypothesis was rejected in nine instances and accepted in nine as regards interaction. Table 2 presents a condensed version of the analysis of variance

---

Insert Table 2 about here

---

in which all non-significant mean squares have been pooled with their respective error terms.

In order to conserve all of the information in the test scores it was necessary to retain scores on the 11 subtests shown in Table 3. Table 3 also shows the contrasts found to be significant, and the mean reliability per item of each subtest.

Insert Table 3 about here

This last statistic is, of course, equivalent to what whould be obtained by using the Spearman-Brown formula backwards on each subtest, "prophesying" the reliability of a one-item test in each instance (Gulliksen, 1950, pp. 77-79). The mean reliability per item gives a pretty good idea of the extent to which each subtest is saturated with its own principal component, and its magnitude is independent of the number of items on the subtest.

The analysis of primary interest to ustoday is the one summarized in Table 4. Each of the 15 behavior dimensions in turn was regressed on the 11 NTE

Insert Table 4 about here

---

subtests. Eight of the equations obtained are shown in the table. Neither the multiple correlation nor any of the beta weights in any of the other seven equations was significantly different from zero, so none of them are shown.

Since only two of the 15 equations resulted in a multiple correlation whose probability under the null hypothesis was less than .05, and since only nine of the 165 beta weights met this criterion, these findings should be regarded as tentative only. Because data of this type are so rare, however, they may be worth peeking at. Attempts to predict teacher competence (as measured by various criteria) from ~~their~~ teachers' scores on cognitive tests have been uniformly unsuccessful in the past (G. Barr, 1948). Here we have asked a different question. Instead of trying to predict some amorphous construct called "competence" we have tried to predict stable patterns of classroom behavior which may be regarded as elements of teacher style, and which are clearly defined in operational terms.

The most impressive finding in Table 4 is the multiple correlations of .66 between Lecturing Behavior (as scored on the Flanders' records) and per-

formance on the NTE. Inspection of the beta weights in the equation indicates that the scores a teacher obtains on science items and on items related to teaching principles and practices are principally responsible for this relationship. Teachers who do better on the science items lecture more; teachers who do better on the teaching principles and practices items lecture less.

Results obtained in the larger study indicate that science teachers as a group tend to lecture more than other teachers, so the contribution of the science subtest to the regression equation may be a function of subject taught, in part at least. However, since there were only seven science teachers among the 53 included in the study, it is likely that teachers of other subjects who had high science subtest scores also tended to act like science teachers no matter what subject they taught.

The negative relationship between lecturing and knowledge of items related to teaching principles and practices is intriguing, suggesting as it does that the teacher who lectures may do so only because he does not know any better way to teach!

In looking at the rest of the results in Table 4 let us remember that they are only suggestive, not conclusive. And in doing so let us ask ourselves the question: if a teacher gets his highest score on items of one particular type, what kind of teaching behavior would you expect him to exhibit in his classroom?

The teacher scoring highest on science items tends to be high on Lecturing Behavior, as we have noted, and also on the Modified Content Cross and the I-D Contrast on Student Responses.

What these dimensions all have in common is a sensitivity to two of Flanders' ten categories: Lecturing and Asking Questions, plus a negative weighting on pupil responses, particularly those to which the teacher reacts in a direct fashion. The teacher whose forte is science spends a lot of time dealing with subject-matter, and tends to dominate the discussion himself.

No patterns emerge for teachers with high scores on Mathematics or English A. Those who score highest on English B, which was designed to measure effectiveness of expression, are low on Listening Behavior. This means

that pupil comments in their classroom tend to be brief; their students do not deliver monologues or speak at length without teacher interruption.

The teacher who scores highest on literature items tends to behave in the opposite fashion. He listens to his pupils more, and lets them speak at length; he seldom finds cause to rebuke them, and his own comments tend to be brief.

Teachers scoring high on social studies and art items show no clear-cut pattern; as far as subject-matter content goes, then, it appears to be science, effectiveness of expression, and literary acquaintance that relate to teaching style.

When we turn to professional knowledge, we find that (as noted) teachers who know most about teaching principles and practices tend to teach by question-and-answer rather than by the lecture method. This is confirmed on both of our Lecturing Behavior measures.

The teacher who scores highest on the 20 items devoted to the history and philosophy of education also scores high on Rebuking Behavior and on the I-D Contrast on Student Response. Such a teacher rebukes pupils frequently

(but without rancor or hostility), and at the same time reacts positively to pupil responses to teacher questions. The suggestion is that such a teacher's classroom has a noisy but positive climate.

The teacher who scores high on "School and Society" items is one who asks questions calling for thoughtful and original answers, and who asks students to evaluate and elaborate their own responses. Since this teacher is also the one who criticizes or rebukes pupils most frequently, this suggests a teacher who challenges pupils with difficult questions.

A more sophisticated analysis of these data--perhaps one using canonical correlations--might have made them appear more clear-cut, but we were reluctant to base any more complex analysis than the one reported on so slight a data base lest we inflate their apparent importance more than we may already have done. Let us conclude by stating two inferences which we feel the data do justify.

First, they strongly suggest that the amount and kind of cognitive equipment a teacher possesses is an important element in determining his teaching style.

Second, there is considerable promise in the methodological strategy used in this study--that is, in trying to relate teacher knowledge to teacher behavior. If we had a hundred teachers or more--instead of 53--the results in Table 4 suggest that we would have learned quite a bit about how to predict teacher behavior from tests administered to them before they began to teach, and that we might learn something about the dynamics and etiology of teaching styles as well.

References

Amidon, E. J., & Flanders, N. The role of the teacher in the classrooms. Minneapolis, Minnesota: Paul S. Amidon and Associates, Inc., 1963 In E. J. Amidon & J. B. Hough (Eds.), Interaction Analysis: Theory, Research, and Application. Reading, Massachusetts: Addison-Wesley, 1967.

Barr, A. S. The measurement and prediction of teaching efficiency: A summary of investigations. Journal of Experimental Education, 1948, 16, 203-283.

Gulliksen, H. Theory of mental tests. New York: Wiley, 1950.

Hoyt, C. J. Test reliability estimated by analyses of variance. Psychometrika, 1941, 6, 153-160.

Medley, D. M., & Hill, R. A. A comparison of two techniques for analyzing classroom behaviors. American Educational Research Association, Chicago, Illinois, 1968. In R. C. Burkhart (Ed.), The Assessment Revolution: New Viewpoints for Teacher Evaluation. Albany, New York State Education Department, Division of Teacher Education and Certification, 1969.